

Express Mail No. EL610089383US

IBM DOCKET: ROC9-2000-0313-US1
WHE DOCKET: IBM-181

APPLICATION
FOR
UNITED STATES LETTERS PATENT

TITLE: AUTOMATED NODE RESTART IN CLUSTERED COMPUTER
SYSTEM

APPLICANTS: Jennifer Anne Dervin, Robert Miller, Vicki Lynn Morey and
Kiswanto Thayib

ASSIGNEE: International Business Machines Corporation

Wood, Herron & Evans, L.L.P.
2700 Carew Tower
Cincinnati, Ohio 45202
513-241-2324

SPECIFICATION

AUTOMATED NODE RESTART IN CLUSTERED COMPUTER SYSTEM

Field of the Invention

The invention is generally directed to clustered computer systems, and in particular, to the handling of node restart operations thereon.

Background of the Invention

“Clustering” generally refers to a computer system organization where multiple computers, or nodes, are networked together to cooperatively perform computer tasks. An important aspect of a computer cluster is that all of the nodes in the cluster present a single system image — that is, from the perspective of a user, the nodes in a cluster appear collectively as a single computer, or entity.

Clustering is often used in relatively large multi-user computer systems where high performance and reliability are of concern. For example, clustering may be used to provide redundancy, or fault tolerance, so that, should any node in a cluster fail, the operations previously performed by that node will be handled by other nodes in the cluster. Clustering is also used to increase overall performance, since multiple nodes can often handle a larger number of tasks in parallel than a single computer otherwise could. Often, load balancing can also be used to ensure that tasks are distributed fairly among nodes to prevent individual nodes from becoming overloaded and therefore maximize overall system performance. One specific application of clustering, for example, is in providing multi-user access to a shared resource such as a database or a storage device, since multiple nodes can handle a comparatively large number of user access requests, and since the shared resource is typically still available to users even upon the failure of any given node in the cluster.

Clusters typically handle computer tasks through the performance of “jobs” or “processes” within individual nodes. In some instances, jobs being performed by

different nodes cooperate with one another to handle a computer task. Such cooperative jobs are typically capable of communicating with one another, and are typically managed in a cluster using a logical entity known as a "group." A group is typically assigned some form of identifier, and each job in the group is tagged with that identifier to indicate its membership in the group.

Member jobs in a group typically communicate with one another using an ordered message-based scheme, where the specific ordering of messages sent between group members is maintained so that every member sees messages sent by other members in the same order as every other member, thus ensuring synchronization between nodes. Requests for operations to be performed by the members of a group are often referred to as "protocols," and it is typically through the use of one or more protocols that tasks are cooperatively performed by the members of a group. One example of a protocol utilized by many clusters is a membership change protocol, which permits member jobs to be added to or removed from a group. Another example of a protocol is a node start protocol, which enables new nodes to be added to a cluster.

Clustered computer systems place a high premium on maximizing system availability. As such, automated error detection and recovery are extremely desirable attributes in such systems. One potential source of errors is that of a node failure, which ultimately requires that a node be expelled from a cluster before the node can resume clustering. For example, in many clustered computer systems, individual nodes rely on an underlying clustering infrastructure, often referred to as clustering resource services. Due to various error conditions, such as the failure of a cluster-critical job, or a failure within the clustering infrastructure, the infrastructure may need to be re-initialized to permit the node to re-register with the other nodes in a cluster.

In most instances, it would be extremely desirable to automatically recover from a node failure and reconnect the node to the cluster. In some instances, a node may lose communication with other nodes in a cluster, whereby extraordinary measures may be required to reconnect a node to a cluster. However, in other instances, a failure on a node (e.g., a failure in a cluster-critical job) may not

immediately affect communications of that node with other nodes in a cluster. In these latter types of failures, a node may lose cluster registration, and appear to other nodes in the cluster that the node is dead. Nonetheless, the node may be functional and alive, but incapable of participating in a cluster. In such instances, it is often desirable to "restart" the node to reintroduce the node to the cluster and re-establish clustering on the node.

As an example, a cluster-wide monitoring job may be used in the various nodes in a cluster to monitor the activities of other member jobs executing on the cluster. If such a monitoring job fails on a node, the node must end, since there is nothing doing the monitoring on that node. Restarting just the monitor may not be sufficient because, while the monitor was down, other jobs the monitor was supposed to monitor may have also gone down. It would also be complicated for a restarted monitor to ascertain what may have happened while the monitor was down.

Conventionally, resolution of the failure of a cluster-critical job requires that the node leave the cluster, and then be restarted to add the node back into the cluster in much the same manner as a node is initially added to a cluster. Typically, the restart of a node is initiated via a manual operation by an administrator or operator, or via an automated script executing on the node. A manual operation necessarily requires human intervention, and thus is prone to human error, as well as reduced system availability while an administrator manually restarts the node.

An automated script running on a failed node is also problematic, since a failed node may be incapable of re-joining a cluster after the node has failed. In particular, a failing node may not be capable of determining what caused its failure. Moreover, if the reason for failure is the loss of clustering information required to join with a cluster (e.g., cluster membership data), the node may not be capable of determining how the node joins with an existing cluster. Furthermore, if the failure that required the node to be restarted was incapable of being remedied through a simple restart procedure, a potential exists that an automated script would lock-up while attempting to continually restart the node without success.

Therefore, a significant need exists in the art for a manner of automating the process of detecting and initiating the restart of a node in a clustered computer system, in particular, to increase system availability and reduce operator intervention.

100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290

Summary of the Invention

The invention addresses these and other problems associated with the prior art in providing an apparatus, program product and method that initiate a restart of a node in a clustered computer system using a member of a cluster group that resides on a different node from that to be restarted.

In various embodiments consistent with the invention, a restart operation may be initiated by the member on the different node in response to a membership change message sent by another group member that is resident on the node to be restarted, with an indicator associated with the membership change message that indicates that a restart should be initiated. Moreover in some embodiments, a restart may be implemented in much the same manner as a start operation that is performed when a node is initially added to a cluster. Additional functionality, however, is typically utilized to preclude repeated restart attempts upon a failure of a restart operation.

By initiating a restart from a node other than a failing node, efficient and reliable node restart operations are typically capable of being performed without operator intervention. Moreover, such operations can typically be initiated even if the failing node has lost any information that would otherwise be required to initiate a restart from that node directly. As such, a clustered computer system incorporating such functionality typically has greater reliability and adaptability, and thus improved system performance over conventional systems.

These and other advantages and features, which characterize the invention, are set forth in the claims annexed hereto and forming a further part hereof. However, for a better understanding of the invention, and of the advantages and objectives attained through its use, reference should be made to the Drawings, and to the accompanying descriptive matter, in which there is described exemplary embodiments of the invention.

Brief Description of the Drawings

FIGURE 1 is a block diagram of a clustered computer system consistent with the invention, illustrating an exemplary clustering failure and automated node restart operation on the clustered computer system.

FIGURE 2 is a block diagram of a node in the clustered computer system of Fig. 1.

FIGURE 3 is a flowchart illustrating the program flow of a clustering failure operation initiated by a cluster control job on a failing node in the clustered computer system of Fig. 1.

FIGURE 4 is a flowchart illustrating the program flow of a process MCM leave message routine performed by a cluster control job on a node in the clustered computer system of Fig. 1.

Detailed Description

The embodiments described hereinafter utilize the cluster in which a failed node participates, and in particular a node other than the failed node, to initiate an automated restart of the failed node. In particular, the embodiments described hereinafter rely on a notification from a group member residing on a failed node to induce another group member that resides on a node different from the failing node to initiate a restart operation on the failed node. The principles of the invention may apply to various clustering environments, where multiple nodes collectively perform tasks, and typically present a single system image to external devices and users. A wide variety of clustering environments and clustering infrastructures may be utilized in connection with the invention.

As shown in Fig. 1, for example, a clustered computer system 8 may include a plurality of nodes 10 interconnected with one another via a network of interconnections 11. Any number of network topologies commonly utilized in clustered computer systems may be used consistent with the invention. Moreover, individual nodes 10 may be physically located in close proximity with other nodes, or may be geographically separated from other nodes, e.g., over a wide area network (WAN), as is well known in the art.

In the context of a clustered computer system, at least some computer tasks are performed cooperatively by multiple nodes executing cooperative computer processes (referred to herein as "jobs") that are capable of communicating with one another. Such cooperative jobs are logically organized into a "group", with each cooperative job being designated as a "member" of the group. Group members, however, need not necessarily operate on a common task -- typically all that is required for members of a group is that such members be capable of communicating with one another during execution.

Fig. 1, for example, illustrates an exemplary cluster of nodes 10, also denoted herein for purposes of example by the sequential identifiers 1, 2, 3 . . . N. Resident within various nodes are a plurality of jobs J1-J6 forming the members of an exemplary group in the clustered computer system. As shown in this figure, nodes in a clustered computer system are not required to participate in all groups (e.g., node 5).

Moreover, multiple jobs from a given group may be resident in the same node (e.g., jobs J1 and J2 in node 1).

Also illustrated in Fig. 1 is an exemplary cluster control group, including a group member, referred to herein as a cluster control (CTL) job, resident on every node participating in the cluster (e.g., jobs CTL1-CTLN, resident respectively on nodes 1-N). Most if not all conventional clustering environments incorporate functionality analogous to a cluster control group, which typically manages various cluster functions such as starting and ending nodes, adding or removing nodes to or from a cluster, etc., and which typically requires that a member job be resident on each node participating in a cluster. Moreover, it is anticipated that each node participating in a cluster will incorporate lower-level functionality, referred to herein as either the clustering infrastructure or clustering resource services, which handles basic clustering functionality, e.g., inter-node communications, message ordering, heartbeat monitoring, etc.

It is anticipated that the cluster control group (or another analogous group) will typically implement much of the automated restart functionality described herein, e.g., through the initiation of messages that are communicated between nodes by the clustering resource services of the various nodes in a cluster. However, it should be appreciated by one of ordinary skill in the art having the benefit of the instant disclosure that the invention does not require implementation of automated restart functionality within a cluster control group, and with the assistance of the clustering infrastructure, in all instances. Therefore, the invention is not limited to the particular clustering environment disclosed herein.

As an example of a node failure, assume that the exemplary group comprising jobs J1-J6 is a cluster-critical group, whereby a member of such group must be active and resident on a node for that node to participate in a cluster. Assume for the purposes of the example that cluster-critical job J4 fails on node 3 of Fig. 1. Consistent with the invention, a failure in job J4 will result in a clustering failure on node 3. In this context, the clustering failure refers to the type of node failure where a node loses cluster registration and appears to be dead from the perspective of the other nodes in the cluster. Nonetheless, a clustering failure of the type described herein

includes a type of failure where the clustering infrastructure on a failing node is still capable of sending messages to the cluster, at least long enough to notify the other nodes in the cluster of a need to restart the node. It will be appreciated that in instances where the clustering infrastructure has a fatal error, and no messages may be sent to other nodes of the cluster, alternative functionality would typically be required to initiate a node restart.

In the illustrated embodiments, the types of errors that may initiate a clustering failure capable of initiating an automated node restart include errors such as the failure of cluster-critical jobs such as monitor jobs, storage access jobs, name server jobs, etc. In some embodiments, a cluster-critical job may also incorporate a cluster control job that participates in a cluster control group as described herein. Other relevant errors may include errors related to the corruption of vital clustering data required by a node, e.g., group or member lists, and other information that a node may require to participate in clustering.

In the illustrated embodiments, various entities within a node may locally detect a clustering failure on the node. For example, a failure may be detected by any member job resident on that node, including, for example, a cluster control job. Detection of errors may also be performed by the clustering infrastructure. Other logical entities capable of detecting errors in a computer system may also be used in the alternative.

To address a detected node failure, embodiments consistent with the invention notify the cluster, and in particular, a group member resident on another node in the cluster, of the failure of the node, with the notified member initiating the restart of the failed node. Typically, the notification relies on the existing ordered messaging system in the cluster. Moreover, as will become more apparent below, the initiated restart operation typically operates in much the same manner as a conventional start operation that is used whenever a new node is being added to a cluster. Furthermore, functionality is incorporated to detect recursive restart attempts, e.g., as may occur if a node failure occurs during a restart operation. The additional functionality therefore prevents repeated, recursive attempts to restart a node, when restart via the mechanism described herein is not possible.

Given that clustered computer systems typically permit inter-node communication predominantly between members of the same group, it is anticipated that in most embodiments, the notification of an error in a failing node, and the initiation of a restart operation, will be performed by different members within the same group. In the illustrated embodiments, for example, such operations are performed by the cluster control members on the failing and another node in the cluster. Other entities, however, may perform such functionality in other embodiments consistent with the invention.

It will be appreciated that nomenclature other than that specifically used herein to describe the handling of computer tasks by a clustered computer system may be used in other environments. Therefore, the invention should not be limited to the particular nomenclature used herein, e.g., as to protocols, requests, messages, jobs, groups, etc.

Now turning to Fig. 2, an exemplary hardware configuration for one of the nodes 10 in clustered computer system 8 is shown. Node 10 generically represents, for example, any of a number of multi-user computers such as a network server, a midrange computer, a mainframe computer, etc. However, it should be appreciated that the invention may be implemented in other computers and data processing systems, e.g., in stand-alone or single-user computers such as workstations, desktop computers, portable computers, and the like, or in other programmable electronic devices (e.g., incorporating embedded controllers and the like).

Node 10 generally includes one or more system processors 12 coupled to a main storage 14 through one or more levels of cache memory disposed within a cache system 16. Furthermore, main storage 14 is coupled to a number of types of external devices via a system input/output (I/O) bus 18 and a plurality of interface devices, e.g., an input/output adaptor 20, a workstation controller 22 and a storage controller 24, which respectively provide external access to one or more external networks (e.g., a cluster network 11), one or more workstations 28, and/or one or more storage devices such as a direct access storage device (DASD) 30. Any number of alternate computer architectures may be used in the alternative.

To implement automated node restart functionality consistent with the invention, each node in a cluster typically includes a clustering infrastructure to manage the clustering-related operations on the node. For example, node 10 is illustrated as having resident in main storage 14 an operating system 30 implementing a cluster infrastructure referred to as clustering resource services 32. One or more jobs or applications 34 are also illustrated, each having access to the clustering functionality implemented within clustering resource services 32. Moreover, node 10 typically includes a cluster control (CTL) job 36 that participates in a cluster control group to assist in managing clustering functionality on behalf of the node. It will be appreciated, however, that the functionality described herein may be implemented in other layers of software in node 10, and that the functionality may be allocated among other programs, computers or components in clustered computer system 8. Therefore, the invention is not limited to the specific software implementation described herein.

The discussion hereinafter will focus on the specific routines utilized to implement the above-described automated node restart functionality. The routines executed to implement the embodiments of the invention, whether implemented as part of an operating system or a specific application, component, program, object, module or sequence of instructions, will also be referred to herein as "computer programs," or simply "programs." The computer programs typically comprise one or more instructions that are resident at various times in various memory and storage devices in a computer, and that, when read and executed by one or more processors in a computer, cause that computer to perform the steps necessary to execute steps or elements embodying the various aspects of the invention. Moreover, while the invention has and hereinafter will be described in the context of fully functioning computers and computer systems, those skilled in the art will appreciate that the various embodiments of the invention are capable of being distributed as a program product in a variety of forms, and that the invention applies equally regardless of the particular type of signal bearing media used to actually carry out the distribution. Examples of signal bearing media include but are not limited to recordable type media such as volatile and nonvolatile memory devices, floppy and other removable disks,

hard disk drives, optical disks (e.g., CD-ROM's, DVD's, etc.), among others, and transmission type media such as digital and analog communication links.

It will be appreciated that various programs described hereinafter may be identified based upon the application for which they are implemented in a specific embodiment of the invention. However, it should be appreciated that any particular program nomenclature that follows is used merely for convenience, and thus the invention should not be limited to use solely in any specific application identified and/or implied by such nomenclature.

Now turning to Figs. 3 and 4, a specific embodiment of the invention, for use in an OS/400 clustering environment utilizing AS/400 or I series servers from International Business Machines Corporation is described. Fig. 3, in particular, illustrates at 50 the program flow associated with handling a clustering failure on a failing node. Fig. 4 illustrates a process Membership Change Message (MCM) leave routine 80 executed by another node in a clustered computer system to initiate a restart operation consistent with the invention.

As will become more apparent below, this embodiment of the invention relies on a membership change protocol to notify other nodes in a cluster of the need to restart a failed node. See, for example U.S. S/N 09/481,177, filed on January 12, 2000 by Funk, Goft, Kloper, Pinter and Yeger-Lotem, which is incorporated by reference herein, for a discussion of membership change protocol and ordered messaging in general. In addition, a start protocol, which operates in the same manner as for adding a new node to a clustering environment, is also used. It will be appreciated, however, that in other clustering environments, e.g., in environments supported by other manufacturers, alternate protocols may be utilized to perform the notification and restart initiation operations described herein. Therefore, the invention is not limited to the OS/400 environment described herein.

In general, an automated restart operation as described in connection with Figs. 3 and 4 begins with the detection of a clustering failure in a node, which requires that the node unregister with the cluster. The cluster control group member resident on the failing node unregisters with the cluster using a Membership Change Message (MCM) to initiate a membership change protocol on the other nodes of the system.

Associated with the MCM is a restart indicator that indicates that the purpose of the membership change is to restart the node. Upon initiation of the membership change protocol, the cluster control job on the failed node ends as in a normal error situation.

Upon receiving the membership change protocol, the remaining members of the cluster control group on the other nodes in the cluster select one member to submit a start node request and initiate a start node protocol on the failed node. In this embodiment, the lowest-named member is selected to submit the start node request, and an indicator associated with the request is used to indicate that the start request is a "restart". The start node protocol creates a new cluster control member on the previously-failed node, and the node completes its join in the same manner as if the node were being newly added to the cluster. If the cluster control member on the previously-failed node tries to unregister during the start protocol, the start message may be checked to determine if the start message indicated a restart. Through checking this condition, the cluster control member can prevent recursive attempts to restart the node.

Turning now to Fig. 3, the principal operations in handling a clustering failure on a failing node are illustrated at 50. This sequence of operations is performed by the clustering control job on the failed node, and may be initiated in response to detection of a clustering failure either by the clustering control job, the clustering resource services, or another job resident on the node. Beginning at block 52, it is determined whether a restart indicator has been set for the node. As will be described in greater detail below, the restart indicator is typically supplied in the initial start message that is sent by another node when attempting to restart a failed node. The restart indicator may be represented using any suitable data structure, and is predominantly used to distinguish the restart operation from a conventional start operation that might be manually initiated by a user on another node.

Assuming first that the restart indicator is not set, control passes to block 54 to unregister the node from the cluster, using a "restart" reason code. Typically, unregistration of the node is performed via an unregister function call 56 to the clustering resource services resident on the failed node.

In the herein-described embodiment, various reason codes are capable of being incorporated into an unregistration call, as well as in a membership change request used to initiate a membership change protocol. To support the automated restart functionality described herein, a dedicated "restart" reason code is used. Any suitable data structure may be utilized to signify a restart condition.

In response to invocation of the unregister function 56 by the cluster control job, the clustering resource services sends a Membership Change Message (MCM) in block 58 with the "restart" reason code (e.g., via setting a reason field in the message to a value associated with a restart) incorporated into the message. Sending of the Membership Change Message notifies the other nodes in the cluster of the need to restart the failed node.

Based upon the ordered messaging utilized in the clustered computer system, clustering resource services will receive a response that indicates that the message was sent and received. A confirmation, designated at 60, is therefore returned to the cluster control job upon completion of the initiation of the membership change protocol. The cluster control job then ends clustering on the node at block 62, typically by invoking an end clustering function 64 to terminate the clustering resource services as shown at block 66. In addition, once termination of the clustering resource services is initiated by the cluster control job, control passes to block 68 to terminate the cluster control job itself. Thus, upon the completion of blocks 66 and 68, all clustering on the node is terminated.

Now turning to Fig. 4, the processing of the Membership Change Message sent to the other nodes in the cluster is illustrated by a process MCM routine 80. A Membership Change Message having a "restart" reason code is considered a type of "leave" membership change request, as it is desired for the node signaling the restart to "leave" the cluster. Routine 80 therefore represents the MCM leave functionality for every node in the cluster, and is typically handled by the cluster control job on that node.

Routine 80 begins in block 82 by determining whether the reason code for the received MCM has a "restart" code. If not, conventional leave processing is performed as shown at block 84, and routine 80 is complete.

However, in the case of a restart reason code, block 82 passes control to block 86 to determine whether the local node is the lowest-named member for the cluster control group. The decision in block 86 essentially assigns one member of the cluster control group to handle the initiation of the restart operation, and thus prevents

multiple restart operations from being initiated by multiple nodes in the cluster. Other methods of selecting one member to initiate the restart operation may be used in the alternative. For example, in some embodiments, one member of a cluster control group may be designated the "leader", whereby any non-leader nodes would simply defer processing of the restart request to the leader.

If the cluster control member executing routine 80 is not the lowest-named member, control passes directly to block 84. On the other hand, if the member is the lowest-named member, and is thus assigned the responsibility of initiating the restart operation, control passes to block 88 to submit a start node request to the group, with a restart indicator set in the request. Submission of the request initiates a start node protocol, which may operate in the same manner as any other start node protocol used to start a node and initialize the node for clustering. The only difference being that a restart indicator is set to ensure that the node being started recognizes that the node is being restarted.

Various start protocols may be utilized consistent with the invention. For example, in the aforementioned OS/400 environment, a start protocol may generally operate by using a dedicated TCP/IP port to the inetd daemon to start a cluster-related job on the node. The job started on the node would then spawn a cluster control member, and forward the message data sent to the inetd port to the cluster control member to initiate the registration of the node with the cluster.

It will be appreciated that, due to ordered messaging, the start protocol will not be processed until the MCM protocol is complete. As such, prior to the start protocol being received by the failed node, both the cluster control member and the clustering infrastructure will typically have been terminated on the node.

It will be appreciated that other protocols for starting a node may be used in the alternative. For example, other TCP/IP ports, or a shared file system, may be used to start a job for initiating clustering on a node. Generally, no modification of the start

protocol, other than the addition of a restart indicator, would be required to support automated restart functionality as described herein.

Returning now to Fig. 3, prevention of recursive attempts to restart a node is supported through detection of the restart indicator in block 52. As described above, the program flow shown at 50 is called in response to a clustering failure. Block 52 therefore determines whether the clustering failure occurred during a restart attempt on the node, indicating potentially that the node is incapable of being restarted in the manner described herein. If the restart indicator is set, block 52 passes control to block 70 to determine whether a protocol count exceeds a predetermined threshold.

The protocol count is a running tally of the number of protocols that is successfully completed by a node since clustering was started on the node. The protocol count value is incremented each time a protocol is processed. The threshold may be empirically determined to ensure that, after a node has been restarted, and has been operating for an adequate period of time, the restart indicator will no longer be capable of preventing the node from being restarted in the future. This accommodates for the fact that a node may be restarted, yet operate long enough to essentially be up and running, whereby a later error may permit another restart operation to occur.

Therefore, if the protocol count is not greater than a predetermined threshold, block 70 passes control to block 72 to signal a failure, indicating that an automated restart operation is not capable of being performed. Otherwise, if the protocol count exceeds the threshold, block 70 may either simply pass control to block 54, or in the alternative, may optionally clear the restart indicator as shown at block 74. Block 74 is optional, given that as long as the protocol count exceeds the threshold, the presence of a set restart indicator will not prevent a restart operation from occurring. Nonetheless, it may be desirable in some embodiments to clear the restart indicator sometime after the node is up and running to permit the node to be restarted at a later time if required.

The signal of a failure in block 72 may result in the initiation of an MCM leave protocol to unregister the node with the cluster. However, in such an instance, a failure reason, rather than a restart reason, would be indicated in the message. A

program flow similar to blocks 54-68 would be performed, but with the reason code set to indicate a failure.

5 Various modifications may be made to the illustrated embodiments without departing from the spirit and scope of the invention. For example, a restart protocol that is separate from the start protocol may be used in some embodiments. However, by utilizing the same protocol for both starts and restarts, system complexity is reduced.

10 Other modifications will be apparent to one of ordinary skill in the art having the benefit of the instant disclosure. Therefore, the invention lies in the claims hereinafter appended.

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207